Reference Solutions of HW2

Please contact lu-c26@mails.tsinghua.edu.cn for any questions or suggestions.

Problem 1 Here $N = 12$ and $N_{\mathrm{t}} = N_{\mathrm{c}} = 6$. All confidence intervals below use the normal approximation with $z_{0.975} \approx 1.96$.

(a) **Slides 1, pages 21 and 38.** Treatment is assigned to classrooms, so the classroom is the randomization unit. Students in the same class share instruction and may influence one another. Such interference may violate SUTVA if individual students are treated as units.

(b) **Slides 2, pages 44–45.** The finite-sample estimand is $\tau_{\mathrm{fs}} = N^{-1} \sum_{i=1}^{N} [Y_i(1) - Y_i(0)]$. Then

$$\hat{\tau}^{\mathrm{dif}} = \bar{Y}_{\mathrm{t}}^{\mathrm{obs}} - \bar{Y}_{\mathrm{c}}^{\mathrm{obs}} = 14.5, \qquad V_W\left(\hat{\tau}^{\mathrm{dif}}\right) = \frac{S_{\mathrm{t}}^2}{N_{\mathrm{t}}} + \frac{S_{\mathrm{c}}^2}{N_{\mathrm{c}}} - \frac{S_{\mathrm{tc}}^2}{N},$$

Neyman's variance estimate is conservative because it omits the nonnegative term S_{tc}^2/N:

$$\hat{V}_{\mathrm{Neyman}} = \frac{s_{\mathrm{t}}^2}{N_{\mathrm{t}}} + \frac{s_{\mathrm{c}}^2}{N_{\mathrm{c}}} = 60.597,$$

$$95\%\ \mathrm{CI}_{\mathrm{Neyman}} = \hat{\tau}^{\mathrm{dif}} \pm z_{0.975}\sqrt{\hat{V}_{\mathrm{Neyman}}} = [-0.757,\ 29.757], \qquad p_{\mathrm{Neyman}} = 0.063.$$

Alternatively, using $\rho_{\mathrm{tc}} \leqslant 1$ gives a tighter variance upper bound, estimated by

$$\hat{V}_{\rho_{\mathrm{tc}}=1} = \frac{s_{\mathrm{t}}^2}{N_{\mathrm{t}}} + \frac{s_{\mathrm{c}}^2}{N_{\mathrm{c}}} - \frac{(s_{\mathrm{t}} - s_{\mathrm{c}})^2}{N} = 60.532,$$

$$95\%\ \mathrm{CI}_{\rho_{\mathrm{tc}}=1} = \hat{\tau}^{\mathrm{dif}} \pm z_{0.975}\sqrt{\hat{V}_{\rho_{\mathrm{tc}}=1}} = [-0.749,\ 29.749], \qquad p_{\rho_{\mathrm{tc}}=1} = 0.062.$$

Both methods give nearly identical results and intervals covering zero. Neither rejects $H_0 : \tau_{\mathrm{fs}} = 0$ at the 5% significance level.

- Python

```python
# HW2_1b
import numpy as np
from scipy import stats
y_t = np.array([57.6, 43.8, 66.5, 70.9, 72.7, 80.8])
y_c = np.array([29.3, 39.3, 51.1, 59.9, 62.2, 63.5])
def neyman(a, b, rho_one=False):
    tau = a.mean() - b.mean()
    variance = a.var(ddof=1) / len(a) + b.var(ddof=1) / len(b)
    if rho_one:
        n_t, n_c = len(a), len(b)
        n = n_t + n_c
        variance -= (a.std(ddof=1) - b.std(ddof=1)) ** 2 / n
    se = np.sqrt(variance)
    ci = tau + stats.norm.ppf(0.975) * se * np.array([-1, 1])
    p = 2 * stats.norm.sf(abs(tau / se))
    print(f"Estimate = {round(tau, 3):g}, variance = {variance:.3f}")
    print(f"95% CI = [{ci[0]:.3f}, {ci[1]:.3f}], p = {p:.3f}")
print("Neyman:")
neyman(y_t, y_c)
print("rho_tc = 1:")
neyman(y_t, y_c, rho_one=True)
```

```
Neyman:
Estimate = 14.5, variance = 60.597
95% CI = [-0.757, 29.757], p = 0.063
rho_tc = 1:
Estimate = 14.5, variance = 60.532
95% CI = [-0.749, 29.749], p = 0.062
```

- R

```r
# HW2_1b
y_t <- c(57.6, 43.8, 66.5, 70.9, 72.7, 80.8)
y_c <- c(29.3, 39.3, 51.1, 59.9, 62.2, 63.5)
neyman <- function(a, b, rho_one = FALSE) {
  tau <- mean(a) - mean(b)
  variance <- var(a) / length(a) + var(b) / length(b)
  if (rho_one) {
    n_t <- length(a)
    n_c <- length(b)
    n <- n_t + n_c
    variance <- variance - (sd(a) - sd(b))^2 / n
  }
  se <- sqrt(variance)
  ci <- tau + qnorm(0.975) * se * c(-1, 1)
  p <- 2 * pnorm(-abs(tau / se))
  cat(sprintf("Estimate = %g, variance = %.3f\n", round(tau, 3), variance))
  cat(sprintf("95%% CI = [%.3f, %.3f], p = %.3f\n", ci[1], ci[2], p))
}
cat("Neyman:\n")
neyman(y_t, y_c)
cat("rho_tc = 1:\n")
neyman(y_t, y_c, rho_one = TRUE)
```

```
Neyman:
Estimate = 14.5, variance = 60.597
95% CI = [-0.757, 29.757], p = 0.063
rho_tc = 1:
Estimate = 14.5, variance = 60.532
95% CI = [-0.749, 29.749], p = 0.062
```

(c) **Slides 2, page 11.** The pretest mean difference is $\bar{X}_{\mathrm{t}} - \bar{X}_{\mathrm{c}} = -0.783$. Under complete randomization, the pretest scores are fixed and all $\binom{N}{N_{\mathrm{t}}} = \binom{12}{6} = 924$ assignments are equally likely. A two-sided balance check gives

$$p = \Pr\left(\left|\bar{X}_{\mathrm{t}} - \bar{X}_{\mathrm{c}}\right| \geqslant \left|\bar{X}_{\mathrm{t}}^{\mathrm{obs}} - \bar{X}_{\mathrm{c}}^{\mathrm{obs}}\right|\right) = \frac{862}{924} \approx 0.933.$$

At the 5% significance level, we cannot reject the null hypothesis of complete randomization, which means that the observed pretest balance is consistent with random assignment. Note that this result alone does not prove that the assignment was randomized.

- Python

```python
# HW2_1c
from itertools import combinations
x_t = np.array([46.6, 63.2, 65.4, 66.0, 73.4, 84.8])
x_c = np.array([50.0, 55.9, 55.8, 69.8, 81.1, 91.5])
x = np.r_[x_t, x_c]
obs = x_t.mean() - x_c.mean()
# For each treated subset, subtract the mean of its complement.
diffs = np.array([x[list(s)].sum() / 3 - x.sum() / 6
                  for s in combinations(range(12), 6)])
extreme = np.count_nonzero(np.abs(diffs) >= abs(obs) - 1e-10)
print(f"Pretest difference = {obs:.3f}")
print(f"Two-sided exact p = {extreme}/{len(diffs)} = {extreme / len(diffs):.3f}")
```

```
Pretest difference = -0.783
Two-sided exact p = 862/924 = 0.933
```

- R

```r
# HW2_1c
x_t <- c(46.6, 63.2, 65.4, 66.0, 73.4, 84.8)
x_c <- c(50.0, 55.9, 55.8, 69.8, 81.1, 91.5)
x <- c(x_t, x_c)
obs <- mean(x_t) - mean(x_c)
# Each column contains one possible treated subset.
subsets <- combn(12, 6)
```

```
diffs <- apply(subsets, 2, function(s) mean(x[s]) - mean(x[-s]))
extreme <- sum(abs(diffs) >= abs(obs) - 1e-10)
cat(sprintf("Pretest difference = %.3f\n", obs))
cat(sprintf("Two-sided exact p = %d/%d = %.3f\n",
            extreme, length(diffs), extreme / length(diffs)))
```

```
Pretest difference = -0.783
Two-sided exact p = 862/924 = 0.933
```

(d) **Slides 2, pages 25 and 45.** Let $D_i(w) = Y_i(w) - X_i$. Since $D_i(1) - D_i(0) = Y_i(1) - Y_i(0)$, the gain-score estimator targets the same τ_{fs}:

$$\hat{\tau}_D = \bar{D}_{\text{t}}^{\text{obs}} - \bar{D}_{\text{c}}^{\text{obs}} = 15.283, \qquad \hat{V}_D = \frac{s_{D,\text{t}}^2}{N_{\text{t}}} + \frac{s_{D,\text{c}}^2}{N_{\text{c}}} = 28.971,$$

$$95\%\ \text{CI}_D = \hat{\tau}_D \pm z_{0.975}\sqrt{\hat{V}_D} = [4.734,\ 25.833], \qquad p_D = 0.005.$$

Compared with (b), the point estimate changes slightly and the estimated variance decreases substantially, leading to rejection of the null hypothesis of no treatment effect at the 5% significance level.

- Python

```
# HW2_1d
neyman(y_t - x_t, y_c - x_c)
```

```
Estimate = 15.283, variance = 28.971
95% CI = [4.734, 25.833], p = 0.005
```

- R

```
# HW2_1d
neyman(y_t - x_t, y_c - x_c)
```

```
Estimate = 15.283, variance = 28.971
95% CI = [4.734, 25.833], p = 0.005
```

(e) **Slides 2, pages 50 and 53.** Now $\tau_{\text{SP}} = E_{\text{SP}}\left[Y(1) - Y(0)\right]$. Under i.i.d. sampling and complete randomization,

$$E\left(\hat{\tau}^{\text{dif}}\right) = \tau_{\text{SP}}, \qquad \text{Var}\left(\hat{\tau}^{\text{dif}}\right) = \frac{\sigma_{\text{t}}^2}{N_{\text{t}}} + \frac{\sigma_{\text{c}}^2}{N_{\text{c}}}.$$

The point estimate and estimated variance are

$$\hat{\tau}_{\text{SP}} = 14.5, \qquad \hat{V}_{\text{SP}} = \frac{s_{\text{t}}^2}{N_{\text{t}}} + \frac{s_{\text{c}}^2}{N_{\text{c}}} = 60.597,$$

$$95\%\ \text{CI}_{\text{SP}} = \hat{\tau}_{\text{SP}} \pm z_{0.975}\sqrt{\hat{V}_{\text{SP}}} = [-0.757,\ 29.757], \quad p_{\text{SP}} = 0.063.$$

The numerical result matches the Neyman result in (b), but the estimand and sources of randomness differ. Here $\hat{V}_{\text{SP}}$ is unbiased for the super-population variance, whereas $\hat{V}_{\text{Neyman}}$ in (b) is generally conservative.

- Python

```
# HW2_1e
# Same computation, now targeting the district super-population ATE.
neyman(y_t, y_c)
```

```
Estimate = 14.5, variance = 60.597
95% CI = [-0.757, 29.757], p = 0.063
```

- R

```
# HW2_1e
# Same computation, now targeting the district super-population ATE.
neyman(y_t, y_c)
```

```
Estimate = 14.5, variance = 60.597
95% CI = [-0.757, 29.757], p = 0.063
```

(f) **Slides 3, pages 14 and 17.** Fit $Y_i^{\text{obs}} = \alpha + \tau W_i + \varepsilon_i$ without pretest adjustment. The OLS estimate of τ_{SP} is

$$\hat{\alpha} = \bar{Y}_{\text{c}}^{\text{obs}} = 50.883, \qquad \hat{\tau}^{\text{ols}} = \hat{\tau}^{\text{dif}} = 14.5.$$

Under equal potential-outcome variances,

$$\hat{\sigma}^2 = \frac{(N_{\text{t}} - 1)\, s_{\text{t}}^2 + (N_{\text{c}} - 1)\, s_{\text{c}}^2}{N - 2} = 181.792, \qquad \hat{V}_{\text{ols}} = \hat{\sigma}^2 \left(\frac{1}{N_{\text{t}}} + \frac{1}{N_{\text{c}}} \right) = 60.597,$$

$$95\%\ \text{CI}_{\text{SP}} = \hat{\tau}^{\text{ols}} \pm z_{0.975} \sqrt{\hat{V}_{\text{ols}}} = [-0.757,\ 29.757].$$

- Python

```python
# HW2_1f
y = np.r_[y_t, y_c]
w = np.r_[np.ones(6), np.zeros(6)]
design = np.column_stack([np.ones(12), w])
coef = np.linalg.lstsq(design, y, rcond=None)[0]
resid = y - design @ coef
sigma2 = (resid @ resid) / (12 - 2)
variance = sigma2 * np.linalg.inv(design.T @ design)[1, 1]
se = np.sqrt(variance)
ci = coef[1] + stats.norm.ppf(0.975) * se * np.array([-1, 1])
print(f"Intercept = {coef[0]:.3f}, treatment coefficient = {round(coef[1], 3):g}")
print(f"Pooled residual variance = {sigma2:.3f}, variance = {variance:.3f}")
print(f"Normal 95% CI = [{ci[0]:.3f}, {ci[1]:.3f}]")
```

```
Intercept = 50.883, treatment coefficient = 14.5
Pooled residual variance = 181.792, variance = 60.597
Normal 95% CI = [-0.757, 29.757]
```

- R

```r
# HW2_1f
y <- c(y_t, y_c)
w <- rep(c(1, 0), each = 6)
fit <- lm(y ~ w)
coef <- coef(fit)
resid <- residuals(fit)
sigma2 <- sum(resid^2) / df.residual(fit)
variance <- vcov(fit)[2, 2]
se <- sqrt(variance)
ci <- coef[2] + qnorm(0.975) * se * c(-1, 1)
cat(sprintf("Intercept = %.3f, treatment coefficient = %g\n", coef[1], round(coef[2], 3)))
cat(sprintf("Pooled residual variance = %.3f, variance = %.3f\n", sigma2, variance))
cat(sprintf("Normal 95%% CI = [%.3f, %.3f]\n", ci[1], ci[2]))
```

```
Intercept = 50.883, treatment coefficient = 14.5
Pooled residual variance = 181.792, variance = 60.597
Normal 95% CI = [-0.757, 29.757]
```

(g) **Slides 3, page 18.** (Optional.) Allowing unequal variances, the estimate of τ_{SP} is unchanged: $\hat{\tau}_{\text{SP}} = \hat{\tau}^{\text{ols}} = 14.5$. Estimate the variances from residuals within each group:

$$\hat{\sigma}_{\text{t}}^2 = \frac{1}{N_{\text{t}} - 1} \sum_{i=1}^{N} W_i \hat{\varepsilon}_i^2 = s_{\text{t}}^2, \qquad \hat{\sigma}_{\text{c}}^2 = \frac{1}{N_{\text{c}} - 1} \sum_{i=1}^{N} (1 - W_i)\, \hat{\varepsilon}_i^2 = s_{\text{c}}^2, \qquad \hat{V}_{\text{ols2}} = \frac{\hat{\sigma}_{\text{t}}^2}{N_{\text{t}}} + \frac{\hat{\sigma}_{\text{c}}^2}{N_{\text{c}}} = 60.597,$$

$$95\%\ \text{CI}_{\text{SP}} = \hat{\tau}^{\text{ols}} \pm z_{0.975} \sqrt{\hat{V}_{\text{ols2}}} = [-0.757,\ 29.757].$$

- Python

```python
# HW2_1g
# Separate residual variance estimates for the two groups.
sigma2_t = np.sum(resid[w == 1] ** 2) / 5
sigma2_c = np.sum(resid[w == 0] ** 2) / 5
variance = sigma2_t / 6 + sigma2_c / 6
se = np.sqrt(variance)
ci = coef[1] + stats.norm.ppf(0.975) * se * np.array([-1, 1])
print(f"Treated variance = {sigma2_t:.3f}, control variance = {sigma2_c:.3f}")
print(f"Variance = {variance:.3f}")
print(f"Normal 95% CI = [{ci[0]:.3f}, {ci[1]:.3f}]")
```

```
Treated variance = 169.862, control variance = 193.722
Variance = 60.597
Normal 95% CI = [-0.757, 29.757]
```

- R

```r
# HW2_1g
# Separate residual variance estimates for the two groups.
sigma2_t <- sum(resid[w == 1]^2) / 5
sigma2_c <- sum(resid[w == 0]^2) / 5
variance <- sigma2_t / 6 + sigma2_c / 6
se <- sqrt(variance)
ci <- coef[2] + qnorm(0.975) * se * c(-1, 1)
cat(sprintf("Treated variance = %.3f, control variance = %.3f\n", sigma2_t, sigma2_c))
cat(sprintf("Variance = %.3f\n", variance))
cat(sprintf("Normal 95%% CI = [%.3f, %.3f]\n", ci[1], ci[2]))
```

```
Treated variance = 169.862, control variance = 193.722
Variance = 60.597
Normal 95% CI = [-0.757, 29.757]
```

(h) **Slides 2, pages 44 and 53; Slides 3, pages 17–18.** Using $\hat{V}_{\text{Neyman}}$ in (b), parts (b), (e), (f), and (g) give the same point estimate, variance estimate, and normal interval here. The alternative in (b) gives a slightly narrower interval. Part (b) targets these 12 classrooms and uses a conservative variance estimate, while (e), (f), and (g) target the district super-population and include sampling randomness, with

$$\hat{V}_{\text{Neyman}} = \hat{V}_{\text{SP}} = \hat{V}_{\text{ols2}} = \frac{s_{\text{t}}^2}{N_{\text{t}}} + \frac{s_{\text{c}}^2}{N_{\text{c}}},$$

even for an unbalanced design. Part (f) additionally assumes equal variances. Here, since $N_{\text{t}} = N_{\text{c}} = N/2$,

$$\hat{V}_{\text{ols}} = \hat{\sigma}^2 \left(\frac{1}{N_{\text{t}}} + \frac{1}{N_{\text{c}}} \right) = \frac{(N/2-1)\left(s_{\text{t}}^2 + s_{\text{c}}^2\right)}{N-2} \cdot \frac{4}{N} = \left(s_{\text{t}}^2 + s_{\text{c}}^2\right) \cdot \frac{2}{N} = \frac{s_{\text{t}}^2}{N_{\text{t}}} + \frac{s_{\text{c}}^2}{N_{\text{c}}},$$

even when the two sample variances differ. This equality need not hold for an unbalanced design.

Problem 2

(1) We use the notation from Problem 1.

(a) **Slides 2, page 31.**

$$\hat{\tau}^{\text{dif}} = \frac{1}{N_{\text{t}}} \sum_{i=1}^{N} W_i Y_i(1) - \frac{1}{N_{\text{c}}} \sum_{i=1}^{N} (1 - W_i)\, Y_i(0).$$

(b) **Slides 2, page 32.** Since $E(W_i) = N_{\text{t}}/N$ and $E(1 - W_i) = N_{\text{c}}/N$,

$$E\left(\hat{\tau}^{\text{dif}}\right) = \sum_{i=1}^{N} \frac{E(W_i)\, Y_i(1)}{N_{\text{t}}} - \sum_{i=1}^{N} \frac{E(1 - W_i)\, Y_i(0)}{N_{\text{c}}} = \frac{1}{N} \sum_{i=1}^{N} [Y_i(1) - Y_i(0)] = \tau_{\text{fs}}.$$

(c) **Slides 2, page 25.**

$$\hat{\tau}_D = \frac{1}{N_\mathrm{t}} \sum_{i=1}^{N} W_i \left[Y_i(1) - X_i \right] - \frac{1}{N_\mathrm{c}} \sum_{i=1}^{N} (1 - W_i) \left[Y_i(0) - X_i \right].$$

(d) Using the same assignment probabilities,

$$\begin{aligned} E\left(\hat{\tau}_D\right) &= \sum_{i=1}^{N} \frac{E\left(W_i\right)\left[Y_i(1) - X_i\right]}{N_\mathrm{t}} - \sum_{i=1}^{N} \frac{E\left(1 - W_i\right)\left[Y_i(0) - X_i\right]}{N_\mathrm{c}} \\ &= \frac{1}{N} \sum_{i=1}^{N} \left\{ \left[Y_i(1) - X_i\right] - \left[Y_i(0) - X_i\right] \right\} = \tau_\mathrm{fs}. \end{aligned}$$

(e) Unbiasedness concerns averages over assignments, not equality for each assignment:

$$\hat{\tau}_D - \hat{\tau}^\mathrm{dif} = -\left(\bar{X}_\mathrm{t} - \bar{X}_\mathrm{c}\right) = 0.783, \qquad E\left(\hat{\tau}_D - \hat{\tau}^\mathrm{dif}\right) = -E\left(\bar{X}_\mathrm{t} - \bar{X}_\mathrm{c}\right) = 0.$$

Thus both are unbiased, although their realized estimates and variances can differ.

(2) Let $q = N_\mathrm{t}/N$. Conditioning on W_i gives

$$\begin{aligned} E_\mathrm{SP}\left[Y_i^\mathrm{obs}\left(X_i - \mu_X\right)\right] &= E\left\{E_\mathrm{SP}\left[Y_i^\mathrm{obs}\left(X_i - \mu_X\right) \mid W_i\right]\right\} \\ &= qE_\mathrm{SP}\left[Y_i(1)\left(X_i - \mu_X\right) \mid W_i = 1\right] + (1-q)E_\mathrm{SP}\left[Y_i(0)\left(X_i - \mu_X\right) \mid W_i = 0\right] \\ &= qE_\mathrm{SP}\left[Y_i(1)\left(X_i - \mu_X\right)\right] + (1-q)E_\mathrm{SP}\left[Y_i(0)\left(X_i - \mu_X\right)\right] \\ &= \frac{N_\mathrm{t}}{N}\mathrm{Cov}_\mathrm{SP}\left[Y_i(1), X_i\right] + \frac{N_\mathrm{c}}{N}\mathrm{Cov}_\mathrm{SP}\left[Y_i(0), X_i\right]. \end{aligned}$$

The third equality uses complete randomization, which makes W_i independent of $[X_i, Y_i(0), Y_i(1)]$.

Problem 3

(a) **Slides 3, page 20.** Substituting $\hat{Y}_i(1) = \bar{Y}_\mathrm{t}^\mathrm{obs}$ and $\hat{Y}_i(0) = \bar{Y}_\mathrm{c}^\mathrm{obs}$ gives

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i &= \frac{1}{N}\sum_{i=1}^{N}\left\{W_i\left[Y_i(1) - \hat{Y}_i(0)\right] + (1 - W_i)\left[\hat{Y}_i(1) - Y_i(0)\right]\right\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[W_i\left(Y_i^\mathrm{obs} - \bar{Y}_\mathrm{c}^\mathrm{obs}\right) + (1 - W_i)\left(\bar{Y}_\mathrm{t}^\mathrm{obs} - Y_i^\mathrm{obs}\right)\right] \\ &= \frac{1}{N}\left[N_\mathrm{t}\bar{Y}_\mathrm{t}^\mathrm{obs} - N_\mathrm{t}\bar{Y}_\mathrm{c}^\mathrm{obs} + N_\mathrm{c}\bar{Y}_\mathrm{t}^\mathrm{obs} - N_\mathrm{c}\bar{Y}_\mathrm{c}^\mathrm{obs}\right] = \bar{Y}_\mathrm{t}^\mathrm{obs} - \bar{Y}_\mathrm{c}^\mathrm{obs} = \hat{\tau}^\mathrm{dif}. \end{aligned}$$

(b) **Slides 3, pages 15 and 21.** Let $q = \Pr\left(W_i = 1\right) \in (0, 1)$. Complete randomization gives

$$E_\mathrm{SP}\left(Y_i^\mathrm{obs} \mid W_i = 1\right) = \mu_\mathrm{t}, \qquad E_\mathrm{SP}\left(Y_i^\mathrm{obs} \mid W_i = 0\right) = \mu_\mathrm{c},$$

$$E_\mathrm{SP}\left(Y_i^\mathrm{obs}\right) = E\left[E_\mathrm{SP}\left(Y_i^\mathrm{obs} \mid W_i\right)\right] = q\mu_\mathrm{t} + (1-q)\mu_\mathrm{c},$$

$$E_\mathrm{SP}\left(W_i Y_i^\mathrm{obs}\right) = E\left[W_i E_\mathrm{SP}\left(Y_i^\mathrm{obs} \mid W_i\right)\right] = q\mu_\mathrm{t}.$$

Using $E\left(W_i\right) = E\left(W_i^2\right) = q$, setting the partial derivatives of $Q\left(\alpha, \tau\right)$ to zero gives

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} &= -2E_\mathrm{SP}\left(Y_i^\mathrm{obs} - \alpha - \tau W_i\right) \quad \Longrightarrow \quad \alpha^* + q\tau^* = q\mu_\mathrm{t} + (1-q)\mu_\mathrm{c}, \\ \frac{\partial Q}{\partial \tau} &= -2E_\mathrm{SP}\left[W_i\left(Y_i^\mathrm{obs} - \alpha - \tau W_i\right)\right] \quad \Longrightarrow \quad q\left(\alpha^* + \tau^*\right) = q\mu_\mathrm{t}. \end{aligned}$$

Solving these equations yields

$$\alpha^* = \mu_\mathrm{c}, \qquad \tau^* = \mu_\mathrm{t} - \mu_\mathrm{c} = \tau_\mathrm{SP}.$$

These are the unique minimizers because

$$Q(\alpha,\tau)-Q(\alpha^*,\tau^*)=(1-q)(\alpha-\mu_{\mathrm{c}})^2+q(\alpha+\tau-\mu_{\mathrm{t}})^2\geqslant 0,$$

with equality only at $(\alpha,\tau)=(\alpha^*,\tau^*)$ when $0<q<1$. Note that if we condition Q on W_i before differentiating, we obtain the intermediate step $\tau^*=E_{\mathrm{SP}}\left(Y_i^{\mathrm{obs}}\mid W_i=1\right)-E_{\mathrm{SP}}\left(Y_i^{\mathrm{obs}}\mid W_i=0\right)$, then complete randomization gives $\tau^*=\tau_{\mathrm{SP}}$.